Response to reviewers' comments on the manuscript "Satellite observations of stratospheric carbonyl fluoride" by Jeremy J. Harrison et al.

We thank the reviewers for their comments. These comments are reproduced below in bold text, followed by our responses.

Referee #1:

General comments

The SLIMCAT run with fixed year 2000 dynamics is extremely helpful to separate the effects of changes in source gases and changes in stratospheric dynamics. It would be very helpful to have also the global mean trend values for this SLIMCAT sensitivity simulation given (couldn't find it) to compare with the global mean trend including inter-annual changes in transport. Do the changes in stratospheric dynamics affect the global mean trend or is the global mean largely insensitive against changes in dynamics?
If we calculate a SLIMCAT trend over all latitudes up to 60 km, the trend is 0.91%/year for both the fixed and full dynamics calculations. Note that the trend described in the paper is a VMR-weighted trend corresponding to the latitude-altitude range of the measurements; this is 0.88%/year for the full dynamics calculation and 0.60%/year for the fixed calculation. Since the fixed run is not physically realistic, the fixed trend was not discussed in the manuscript.

Does the agreement of trends between MIPAS and SLIMCAT suggest any conclusion whether this trend is consistent with trends in surface mixing ratios of the source gases?
We do not think so. The surface trends in the assumed sources of COF_2 (e.g. CFC-12) are very well known by in-situ surface monitoring. A more important question here would be whether the comparison of the trend can provide any additional information on the accuracy of the model COF_2 chemistry scheme beyond the comparisons shown in the earlier plots. We do not think so because Figure 14 shows the important role of atmospheric dynamics in determining the trends, and this will dominate the comparison.

I suggest to cite and discuss also Kellmann et al. (2012), who found similar trend patterns in MIPAS observations of CFC-11 and CFC-12.
Done.

Could the change from high resolution to low resolution mode in MIPAS observations from 2005 on affect trends in COF2? Was this taken into account in the trend calculation? E.g. by allowing for a possible bias between high and low resolution retrievals?
We are familiar with the work of Kellmann et al., and are aware that they observed a bias in CFC-11 and CFC-12 VMRs between the two MIPAS spectral resolution modes. However, there was no evidence for a similar bias in the COF_2 dataset.

Specific comments
p.18128, l.8: I don't think it is appropriate to count COF2 as "inorganic" fluorine. Better say product gas. However, on p.18129, l.9, "inorganic" is okay and does not need to be written in quotes.

It is accepted in the literature that COF_2 is an inorganic fluorine reservoir. We follow this convention.

p.18135: There is some overlap between Sections 3.1 and 2.1/2.2. I suggest moving the first two paragraphs of Section 3.1 to Section 2.
In order to discuss COF_2 spectroscopic errors, a certain amount of detail needs to be presented in this section. We believe some overlap is acceptable, but have rewritten the beginning of Section 3.1.

p.18140, l.17: The smoothing error covariance is not the same as the a priori error covariance! (The smoothing error covariance is usually defined as S_smooth=(1-A)S_a(1-A)ˆT.) Please rephrase to avoid confusion.
The meaning was that the contribution of the a priori error to the retrieval error is sometimes known as the smoothing error. This has now been reworded.

p.18145, l.18: What is the meaning of "ERA-Interim from 1989 onwards" if SLIMCAT was run from 2000 to 2012? I believe this sentence has been "recycled" and is not relevant here, but if not, then why was ERA-Interim not used before 1989?
The SLIMCAT run was originally spun up from 1977. ECMWF have two reanalysis products that we can use to force the model: ERA-40 (available from 1957-2002) and ERA-Interim. ERA-Interim was originally produced from 1989, but is now available back to 1979. The sentence means that the spin-up run use ERA-Interim from 1989 onwards.

p.18151, l.3: I may have overlooked this, but why is September 2010 the last month for which ACE v3.0 is usable?
All ACE v3.0 processed data from October 2010 suffer from problems in the P and T supplied by the Canadian Meteorological Centre. Version 3.5, which includes correct P and T, should be used for any measurements taken from October 2010. Unfortunately, processing has been slow and v3.5 wasn't ready in time for this study.

p.18152, l.25: why were no errors calculated for the SLIMCAT trends? Although it will not be straight-forward to estimate a "model error", it should be possible to estimate the linear regression error for SLIMCAT in the same way as for the satellite data.
The linear regression error for SLIMCAT trends is much smaller than the ACE/MIPAS errors (which also include observation errors), so will not provide a true comparison.

p.18154, l.10: "see their Fig.10" as this refers to Fig.10 of Stiller et al.
Done.

Reference
Kellmann, S., von Clarmann, T., Stiller, G. P., Eckert, E., Glatthor, N., Höpfner, M., Kiefer, M., Orphal, J., Funke, B., Grabowski, U., Linden, A., Dutton, G. S., and Elkins, J. W.: Global CFC-11 (CCl3F) and CFC-12 (CCl2F2) measurements with the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS): retrieval, climatologies and trends, Atmos. Chem. Phys., 12, 11857-11875, doi:10.5194/acp-12-11857-2012, 2012.
The reference has been added.

Referee #2:

General Comments

1) I would recommend adding a paragraph in the introduction that provides the scientific motivation for measurements and modelling of COF2 from a broader perspective. Why are global COF2 measurements needed? Are they mainly needed to validate and improve modelling of COF2 in the stratosphere? Do atmospheric models such as SLIMCAT have specific deficiencies representing the photochemistry of COF2 or its source gases? Or are the measurements particularly useful to validate the representation of transport and mixing in the stratosphere? Are there other scientific motivations to measure and model COF2?

Monitoring COF_2 as part of the atmospheric fluorine family is important to close the fluorine budget, particularly as the majority of atmospheric fluorine arises from anthropogenic emissions. Models have not been tested against COF_2 observations in detail before. Many standard stratospheric models do not include fluorine chemistry, and the few comparisons that have been done have used ground-based column data. Therefore, there is a need to test how well we understand COF_2 chemistry through model-data comparisons. A paragraph summarising these points will be added to the final ACP paper.

2) From the perspective of a retrieval scientist I would like to ask for a more detailed description of the vertical resolution and sampling of the ACE-FTS and the SLIMCAT data sets. For MIPAS the averaging kernels and estimates of the vertical resolution are presented, which is fine. For ACE-FTS it would be good to provide these information in section 3.2, for example. I also got the impression that the averaging kernels (or at least some estimates of vertical resolution) are not considered in the comparison of the different data sets (section 4-6)? This might be okay, if differences in vertical resolution are not too large, but this should be made more clear in the paper, I think.

The SLIMCAT dataset has a resolution of 1.5-2.0 km; this is now stated in the text.

For the ACE-FTS, the vertical resolution is defined by the sampling unless the separation between measurements is less than the extent of the field-of-view, in which case the vertical resolution is limited to 2-3 km. Although there is some variation in vertical resolution with the beta angle of the measurement, it is often the case that the vertical resolution at high altitudes (above ~ 40 km) is limited by the sampling, while at low altitudes it is limited by the field of view.

As the reviewer states, the differences in vertical resolution between the datasets are not too large so we do not explicitly account for them in the comparisons.

We have incorporated these points into the manuscript.

Specific Comments

p18129, l5-9: Perhaps add a sentence explaining the distinction between "organic" and "inorganic" forms and why this is important here?

Most of the F source gases are organic, but all the atmospheric degradation products are inorganic. We believe this has been covered adequately in the manuscript.

p18130, l25-p18131, l10: Suggest to add references to the individual sections of the paper in this paragraph (which is describing the structure of the paper).

Done.

p18131, l20: Does "vertical resolution" refer to the vertical sampling of the ACE-FTS measurements or the vertical resolution of the COF2 retrieval?

This term refers to the typical vertical resolution of ACE profiles.

p18131, l25-26: "ACE-FTS records [...] spectra over a large portion of the globe" sound a bit vague, I think.
The reference at the end of this sentence provides more details on the ACE-FTS latitude sampling, which is strongly dependent on the SCISAT orbit.

p18132, l18-20: It would be good to mention the vertical sampling (1km?) of the retrieval grid here.
The retrieved VMRs are interpolated onto a 1 km altitude grid.

p18134, l9: Suggest to replace "error covariance" by "covariance", as S_a refers to the variability of the atmospheric state (in optimal estimation theory).
We consider optimal estimation to be the weighting of two independent estimates scaled by their inverse errors, so although S_a represents the covariance of atmospheric variability about the mean state, in this particular context of optimal estimation it seems appropriate to describe it as representing an error covariance.

p18135, l19-21: This paragraph is too short. It provides a detail on retrieval grids (related to interpolation of data) in the middle of other information. Perhaps this information should be given earlier, e.g. somewhere near the description MIPAS tangent heights and retrieval grid?
Done.

p18138, l5-6: "MkIV and ACE-FTS v2.2 profiles from 2004 and 2005 agree well within the measurement errors." Which relative or absolute difference would that be?
The mean VMRs agree within 10%.

p18139, l1-17: Did you calculate the median absolute deviation or the standard deviation of the individual fractional retrieval errors to get the statistics?
Neither. Individual fractional errors for a subset of occultations were simply averaged to provide an overall estimate of the fractional error.

p18139, l18-p18140, l8: I found this description of the approach to handle the top column data in the ACE-FTS retrieval a bit long and confusing. I understood that an a priori profile from ATMOS was used above a latitude-dependent upper boundary, which was scaled by a factor that was retrieved for each profile. However, this approach causes errors, which are currently not considered in your systematic error budget, right? Please clarify.
The referee's summary of the paragraph is correct. It might be better to say that the procedure, rather than the description, is confusing. We are satisfied that most readers will follow the description.

p18141, l2: The term "retrieval gain matrix" is used, but it was not explained before. It may be introduced with Eq. (5) and could be used to simplify Eq. (6).
To define the error analysis applied to MIPAS spectra mathematically would take several more equations, particularly since it also includes the error propagation via the pT retrieval, which would unbalance the whole paper. However to avoid introducing the undefined term 'gain matrix', we have replaced it with 'the linearised form of Eq.(2)'.

p18141, l6-8: The term "Spectroscopic errors in interfering molecules" confuses me. Do you refer to all the spectroscopic errors from trace gases other than COF2? However, at the end of the sentence you are referring to COF2?
For simplicity 'spectroscopic errors' are represented by a single error spectrum representing a 1-sigma perturbation for all major molecules, as well as the target molecule. This sentence has now been reworded.

p18141, l11-14: This is another very short paragraph referring to the measurement and retrieval grids somewhere in the middle of other information. I think it would be better to collect these information in fewer places. Please try to sort this out.
A modified version of the first sentence has been moved earlier, and the second sentence, which discusses errors, has been merged with the previous paragraph.

p18141, l15: I would not say that the averaging kernel matrix describes the "quality" of a retrieval. Perhaps "sensitivity" (to measurement information and a priori information) is a better word?
We have changed the wording to “The sensitivity of the MIPAS COF2 retrieval to the true state can be measured ...”

p18141, l22-23: Why do you show polar profiles for spring equinox? Wouldn’t it be more reasonable to show polar profiles for January or July, representing the different atmospheric conditions of polar summer and polar winter more clearly?
New representative profiles have been plotted for polar winter/summer conditions (December solstice), in addition to midlatitude and equatorial.

p18142, l13-15: "... near-coincident profiles are expected to be similar but not necessarily identical" is a bit vague. Perhaps clarify by saying "... are expected to agree within the error budget and vertical resolution (AVKs) of the retrievals"?
We have changed the text to “...should agree within measurement error, unless there is significant atmospheric variability.” Note that ACE retrievals do not involve averaging kernels.

p18142, l18-19: Are these error bars representing the full (systematic) error budget or just the retrieval covariance?
The error bars are just the retrieval covariance (random error). The text has been changed to reflect this.

p18142, l21-23: You could convolve the ACE-FTS profile with the MIPAS AVKs to test this hypothesis. Further, I noticed that Fig. 3 (bottom, left) shows a dip in the MIPAS profile which is not present in the ACE-FTS profile. What is the reason for that?
For one small figure, this is probably overkill. The cause of the dip in the MIPAS profile in Fig. 7 is not readily obvious, but probably results from minor variations in the COF_2 atmospheric concentrations.

p18142, l24-p18143, l1: The approach for outlier filtering for the ACE-FTS data remains unclear. How are "significant outliers" defined?
Medians of the data were taken and those VMRs outside ~ ±8 MAD (median absolute deviation) were discarded.

p18143, l8-9: Would this diurnal variation be expected based on COF2 photochemistry? (If this is relevant I think the solar local times of the ascending and descending Envisat orbits should be mentioned in the paper.)
As stated in the paper, we calculate the atmospheric lifetime of COF_2 to be fairly long (3.8 years). Therefore, in the region where COF_2 is abundant we would not expect it to have a diurnal cycle. Note also that the slow formation and loss processes only occur in the daytime anyway.

p18143, l19-20: A forward reference to section 5 of the paper would be good. (I was first tempted to think the secondary maximum is a measurement artifact, but you found that nice explanation that it is caused by the HCFC-22 source.)
Done.

p18144, l3-6: The ACE-FTS zonal mean cross-sections appear noisy also at midlatitudes (e.g. 10/2009 or 03/2010), and not just in the tropics, in particular?
The noisy appearance is a consequence of the poor ACE-FTS sampling. ACE sampling is always poor over the tropics, however for a given month coverage at higher latitudes is variable and depends on the satellite orbit. As the words “particularly in the tropics” have caused confusion, they have been removed.

p18144, l14-16: I would first like to clarify that you indeed plotted the "standard error of the mean" (sigma/sqrt(n)) and not the standard deviations (sigma) of the data? I think the calculation of the standard error of the mean is based on the assumption of independent data? Is this applicable here, where individual profiles are likely correlated due to atmospheric chemistry and dynamics? I was wondering if it wouldn’t make sense to show standard deviations instead of standard errors to illustrate atmospheric variability?
Yes, the standard error of the mean (SEM) is plotted. One benefit of calculating the SEM is that the random errors are reduced by 1/sqrt(n). This is particularly important for the ACE-FTS dataset due to the low sampling. It has been shown by Toohey and von Clarmann (AMT, 6, 937-948) that generally the ‘classic’ SEM (assuming strictly independent data) provides a conservative estimate of the ‘empirical’ SEM. For cases in which the random measurement error is larger than the natural variability, the ‘classic’ SEM should be a reasonably accurate estimate.

p18144, l18-19: It seems the ACE-FTS error bars are largest at upper altitudes, but not necessarily at the upper boundary of the altitude range of the retrieval? Is this due to the upper boundary of the retrieval varying with latitude?
This question has led us to find a bug in the IDL code when calculating the SEM near the “boundary”.
Near the boundary some averaged VMRs are derived from both retrieved and scaled *a priori* VMRs. The ACE v3.0 data product does not provide values of sigma for the latter, so calculating the SEM is not trivial. We approximate the SEM by only using contributing VMRs which are actually retrieved, and taking n as the total number of such VMRs. As seen in the revised figure, this provides good agreement compared with lower altitudes.

p18144, l18-29: I found it a bit difficult to identify the different altitude ranges you are referring to in this paragraph. There are the "highest altitudes" and the "very highest altitudes" and several times you refer to "at these altitudes". Perhaps give some numbers to clarify?

We have rewritten the text to clarify this.

p18144, l21-22: It seems not all of the ACE-FTS mean profiles are flat? For instance, the profile at 25-30S has a distinct dip at 35-38 km, which is not captured by MIPAS. I was wondering if this dip in the ACE-FTS profiles is related to the approach used to handle the top column data in the retrieval? (If the top column is over-estimated for some reason, the retrieval would likely compensate by underestimating the COF2 vmrs just below the top column?)
The sentence in the text about the rather flat appearance of ACE profiles, which the reviewer objected to, related only to the tropics. For the August 2010 25-30S plot in Figure 9, it is not so much the dip, but the sudden increase at the top of the retrieved altitude range that most likely results from the approach used. We have now made this additional point in the text.
Note that the latitude bins in this figure were accidentally shifted by 2.5 degrees from what we thought. This has now been fixed.

p18145, l2-4: Would be good to mention the actual number of profiles in the averages (e.g. in the plot title or the figure caption).
We don't feel this will add anything of great note to the manuscript.

p18145, l15-16: What is the vertical grid spacing of the model/simulation in the stratosphere? Is it comparable to ACE-FTS and MIPAS?
It is stated in the text that there are 32 levels from the surface to 60 km. The levels are not evenly spaced in altitude, but the resolution in the stratosphere is around 1.5-2.0 km. This has been added to the manuscript.

p18146, l22-24: What are the relative contributions of the photolytic and O(1D) loss mechanisms?
In order to keep Figure 10 simple, we did not separate the loss rates into separate panels. Inspection of preliminary plots showed that photolysis was by far the major sink and so we noted that in the text and just showed the total loss rate. We have now run the model again and inserted the relative contributions into the text.

p18147, l3-4: It is not clear to me why the analogy between COF2 and NOy is important or relevant here?
There have not been many studies on atmospheric COF_2 and most readers will not be familiar with its distribution in the stratosphere. Our aim was to make an analogy with a much more common and well understood stratospheric tracer. However, this has now been removed.

p18147, l4-6: Is it important to mention the compact correlations between COF2 and CFC-12? (The tracer-tracer method was not used here to estimate the COF2 lifetime.)
See comment above – we just wanted to inform readers of the properties of COF_2 (like NO_y, it has a straight-line correlation with its main source). However, this has now been removed.

p18147, l19-21: I found it surprising that the ACE-FTS two-year means (Fig. 11) show the same noise-like features as the one-year data (Fig. 8) in some cases. For example, in both figures there are noise-like structures/oscillations at 10-30N and 30-45km in June. (Are these features related to individual outliers in the ACE-FTS data rather than noise?)
The ACE-FTS latitude sampling does not repeat every twelve months. Therefore, some latitude bands in the plot will be identical to those in Fig. 8, some latitude bands will be new

(only sampled during the other year), and some latitude bands will have reduced noise (when they are sampled for each of the two years). We have reworded the sentence.

p18148, l1-12: Following the order of the figures it may make more sense to shift this paragraph (which is discussing Fig. 9) before the preceding paragraph (which is discussing Fig. 11)?
Figure 9 was first mentioned in the previous section. The text here is simply revisiting the figure and focusing on the SLIMCAT calculations. As in the previous section, we focus on the zonal plots before the profiles. We don't believe there is any need to shift the text.

p18149, l1-22: A separate figure or additional plots for SH/NH polar summer in Fig. 10 may help to illustrate that the HCFC-22 source is causing the secondary polar summer maximum.
We did consider including an additional figure whilst preparing the initial manuscript, but in the end decided that this would not greatly add to the discussion presented in the text.

p18151, l8-10: I recall that the MIPAS data coverage also varies with latitude, with maxima occurring at high latitudes?
MIPAS is sampled uniformly in time throughout each orbit, which approximately equates to a constant number of profiles in each latitude band. The only latitude-dependence was in an altitude offset of +/- a few km intended to track the tropopause and prevent large numbers of spectra being obscured by cloud, but this does not affect the geographical profile distribution.

p18153, l2-4: It seems the ACE-FTS trends are more noisy than the MIPAS and SLIMCAT trends? Please comment.
Again this is primarily a result of the limited coverage of ACE-FTS measurements over all latitudes.

p18155, l21-22: I think the 30% high bias of MIPAS needs some more discussion. Is it related to uncertainties of the spectroscopic data (noting that ACE-FTS and MIPAS retrieval use rather different microwindows for the retrievals)? Or is the bias related to the different vertical resolutions and/or a priori influences of the different retrievals? Or are there other reasons?
We believe the main reason is the large spectroscopic errors for COF_2 and the use of different microwindows. We have added this to the text.

Table 1 and Table 3: The ACE-FTS and MIPAS retrievals use rather different microwindows. Except for one window near 1234/cm all other microwindows are different. I found that a bit surprising. Please comment.
The retrievals were performed independently by two different teams. MIPAS has superior signal-to-noise than ACE in the region covering the strong 794 cm^{-1} band.

Fig. 2: How many spectra have been averaged?
3547. We have added this to the caption.

Fig. 3: The retrieval error due to spectroscopic errors is rather small. I thought it would be about 15-20% according to the discussion presented in section 3.1? I guess "day" in the plot title refers to "mid-latitude day-time"?
The plot has now been corrected. The plot caption has been corrected to mention that this is indeed for mid-latitude day-time conditions.

Fig. 4-6: All the curves are rather thin and the colors are hard to distinguish on my printed version of the paper. "DFS" in the top left plot is clear (although not explained in the paper), but what is "INF" referring to? Do you show the median profile to illustrate that the selected profile is close to it and that its diagnostics can be considered to be representative for a large fraction of profiles? The retrieval error at 80S (Fig. 4, top, right) is at 8-10%, whereas Fig. 3 shows that the estimated retrieval noise is >15%. I guess this is due to the different atmospheric conditions (mid-latitudes vs. Spring equinox)? The integral of the AVKs should be displayed in the response plot (bottom, left) to indicate the fraction of measurement information in the retrievals. Fix plot title in vertical resolution plots (bottom, right).
Data in figs 4-6 have been rearranged and replotted.
DFS and INF are explained in the new figure caption (INF is Shannon Information Content).
The median was indeed plotted to show that the selected profiles were typical of the latitude band, but has now been removed to improve clarity of the new plots.
The random error, expressed as a percentage of the VMR, is dependent on concentration and, for limb emissions sensors, tangent point temperature, so there is significant variation.
The error budget was for mid-latitude day-time conditions.
The integral of the averaging kernels is now also plotted.
The vertical resolution plot title has been fixed.

Fig. 9: It is a bit difficult to see the mean ACE-FTS curves due to the broad error bars.
The broad ACE-FTS error bars were chosen so that the MIPAS error bars could be easily distinguished on top. We have made the ACE error bars slightly less broad.

Technical Corrections
p18128, l18: spell out SLIMCAT acronym.
SLIMCAT is not an acronym, it is just a name.

p18128, l25: suggest to use brackets for trends, i.e. "(0.85+/-0.34) % yearˆ-1" rather than "0.85+/-0.34 % yearˆ-1" (also in other places below)
We believe that brackets are unnecessarily confusing.

p18130, l26: "...and the investigation more fully of COF2 trends," may need to be reworded (?)
We are happy with the wording.

p18131, l3: spell out SCISAT acronym
SCIentific SATellite

p18137, l10: "Q branch" rather than "Q-branch" (to be consistent with other places)
Done.

p18137, l20: there is no "on the one hand..." before "on the other hand..." (?)
We have removed “on the other hand”.

p18141, l29: suggest to use "\Delta z_i" or "dz_i" as symbol for grid spacing
Done.

p18143, l24: suggest to replace "very low" by "low"
Done.

p18146, l16: write "... _at_~20-40 km" (?)
We don’t believe this is necessary.

p18147, l21: perhaps "averages of monthly averages" or just "averages" would be more clear?
Done.

p18151, l26: write "_which is_ stronger..." (?)
Done.

p18152, l21: "that" instead of "which" (?)
Done.